S-96,583

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Applicants:	Thomas C. Terwilliger	Docket No.:	S-96,583
Serial No.:	10/017,643	Examiner:	A. Marschel
Filed :	December 12, 2001	Art Unit:	1631
For :	METHOD FOR REMOVING ATOMIC-MODEL BIAS IN MACROMOLECULAR CRYSTALLOGRAPHY		

RESPONSE TO NOTIFICATION OF NON-COMPLIANCE WITH 37 CFR 92(c)

The Examiner has issued on August 25, 2004 a notice of non-compliance with 37 CFR 192(c) for the appeal brief filed June 4, 2004. The Examiner has required that the appeal brief contain a reference to U.S. Patent Application S.N. 09/512,962 as a related case and that the Summary of the Invention set out a reference in the specification and figures applicable to each limitation in the claims on appeal.

CERTIFICATE OF MAILING/TRANSMISSION (37 CFR 1.8(a))

I hereby certify that this correspondence is, on the date shown below, being:

MAILING

☐ deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to the: Commissioner for Patents, PO Box 1450, Alexandria, VA 22313-1450.

Date 9-13-04

FACSIMILE

☐ transmitted by facsimile to the United States Patent and Trademark Office

Signature

Ray G. Wilson
(type or print name of person certifying)

Without concurring that the Examiner's requirements comply with MPEP 1206, applicant has amended the appeal brief to include the copending application as a related case and has replaced the concise summary of the invention with a table containing a listing of the claims on appeal with supporting references to the specification and figures in this case in order to meet the Examiner's requirements.

This case is a continuation-in-part of U.S. Patent Applications S.N. 09/769,612 (now U.S. Patent 6,721,664, issued April 13, 1994) and S.N. 09/512,962, which were incorporated by reference and made a part of the disclosure herein. Copies of these cases are incorporated in the Appendices of the revised appeal brief and reference is made to these cases as needed to support the pending claim limitations.

The amended appeal brief is submitted herewith in triplicate.

Respectfully submitted,

Date: 9-13-04

Signature of Attorney

Reg. No. 28,351
Phone (505) 665-3112

Ray G. Wilson
Los Alamos National Laboratory
LC/IP, MS A187
Los Alamos, New Mexico 87545

AF/1631
I Fw

01/30/04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Appellants:	Thomas C. Terwilliger	Docket No.:	S-96,583
Serial No.:	10/017,643	Examiner:	Marschel
Filed :	December 12, 2001	Art Unit:	1631
For :	METHOD FOR REMOVING ATOMIC-MODEL BIAS IN MACROMOLECULAR CRYSTALLOGRAPHY		

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

TABLE OF CONTENTS

Statement of the Real Party in Interest 1

Related Appeals and Interferences 1

Status of All Claims. 1

Status of Amendments. 2

Summary of the Invention. 2

Issue Presented for Review 5

Grouping of the Claims. 5

Argument. 5

Conclusion. 11

Appendix A - Claims on Appeal
Appendix B - Claim Comparison with Rejection
Appendix C - U.S. Patent Application S.N. 09/512,962
Appendix D - U.S. Patent 6,721,664

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Appellants:	Thomas C. Terwilliger	Docket No.:	S-96,583
Serial No.:	10/017,643	Examiner:	Marschel
Filed :	December 12, 2001	Art Unit:	1631
For :	METHOD FOR REMOVING ATOMIC-MODEL BIAS IN MACROMOLECULAR CRYSTALLOGRAPHY		

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

STATEMENT OF THE REAL PARTY IN INTEREST

The Regents of the University of California is the assignee of all right, title, and interest in U.S. Patent Application Serial No. 10/017,643 from the Government of the United States, United States Department of Energy.

RELATED APPEALS AND INTERFERENCES

There is an appeal pending in U.S. Patent Application S.N. 09/512,962, from which the present case is a continuation-in-part.

STATUS OF ALL CLAIMS

This is an appeal from the final rejection (Examiner's Action dated February 24, 2004) of Claims 1-8 currently pending in the subject patent application. No claims have been allowed.

STATUS OF AMENDMENTS

No amendments have been filed subsequent to this appeal.

SUMMARY OF THE INVENTION

The following table provides a reference to specification locations that support the recited claim limitations. U.S. Patent Application 09/512,962 and U.S. Patent Application 09/769,612 (now U.S. Patent 6,721,664) are incorporated by reference into the present application and specification references to these cases are noted below by "'962" and "'664", respectively.

Claim Limitation	Support Location
1. A method for improving an electron density map representing a crystal structure comprising:	p. 1, l. 14-16; p. 5, l. 14-18
(a) obtaining by x-ray diffraction observed structure factor amplitudes for a plurality of reflections from the crystal structure;	p. 1, l. 19-23; p. 5, l. 18-19; p. 7, l. 6-7; p. 9, l. 9-10; '962: p. 1, l. 16-22; p. 16, l. 24-26
(b) selecting a starting set of crystallographic phases to combine with the observed structure factor amplitudes to form a first set of structure factors;	p. 7, l. 5-8 '664: Col. 10, l. 57-65
(c) deriving a first electron density map from the first set of structure factors;	p. 7, l. 5-10 '962: p. 16, l. 25-26
(d) identifying features of the first electron density map to obtain expected distributions of electron density;	p. 7, l. 10-11; p. 9, l. 13-28; p. 10, l. 6-15 '664: Col. 8, l. 62-67; Col. 9, l. 1-4
(e) making a comparison	P. 7, l. 12-14; p. 9, l. 7-19, p. 10, l. 21-27; p. 11, l. 1-11

between the first electron density map and the expected distribution of electron density;	
(f) estimating how changes in the crystallographic phase of a reflection *k* affect the comparison;	p. 7, l. 20-25; p. 8, l. 6-10, l. 19-23 '664: Col. 8, l. 18-32
(g) establishing crystallographic phase probability distributions from the comparisons for the possible crystallographic phases of reflection *k*;	p. 7, l. 26-28; p. 11, l. 26-32; p. 12, l. 1-14
(h) repeating steps (c) through (g) as *k* is indexed through all of the plurality of reflections;	p. 5, l. 27-30; p. 9, l. 7-10; p. 11, l. 30-32; p. 12, l. 1-14
(i) deriving an updated electron density map using crystallographic phases determined to be most probable from the crystallographic phase probability distributions for each one of the reflections;	p. 8, l. 10-12
(j) repeating steps (d) through (i) to obtain a final set of crystallographic phases with minimum bias from known electron density maps; and	p. 8, l. 12-18; p. 12, l. 15-28; p. 13, l. 13-18
(k) forming a final electron density map using the final set of crystallographic phases.	p. 8, l. 10-12; p. 13, l. 8-10; p. 17, l. 19-23
2. The method of Claim 1, wherein identifying features of the electron density map includes making probability estimates of whether each point in the	p. 7, l. 11-14; p. 10, l. 6-15

map is located in a solvent region or a crystal structure region.	
3. The method of Claim 1, wherein identifying features of the election density map includes estimates of whether the electron density at each point in the map is related by non-crystallographic symmetry to electron density at another point in the map.	p. 9, l. 20-28
4. The method of Claim 1, includes estimates of whether a structural motif is located at each point in the map.	p. 3, l. 27-29; p. 4, l. 1-6; p. 9, l. 24-28 '664: Col. 11, l. 64-66
5. The method of Claim 4, wherein the structural motif is a helix.	'664: Col. 11, l. 64-66
6. The method of any one of Claims 1, 2, 3, or 4, wherein the crystallographic phase probability distributions are log-likelihood functions.	p. 7, l. 14-20; p. 9, l. 29-31; p. 10, l. 1-15
7. The method of Claim 1, further including the steps of calculating first and second derivatives for the crystallographic phase probability distributions with respect to the structure factors; and	p. 7, l. 20-23
applying an FFT-based algorithm to determine the most probable crystallographic phase probability distributions.	p. 7, l. 21; p. 12, l. 10-13
8. The method of Claim 1, wherein the step of selecting a starting set of crystallographic phases includes;	p. 14, l. 24-26 '664: Col. 10, l. 57-67

selecting a model crystal structure having similarities to the crystal structure being examined;	
assigning a low weighting factor to structure factors of the model crystal structure; and	p. 19, l. 8-11, p. 20, l. 3-7
combining the weighted structure factors with the observed structure factors for deriving the first electron density map.	p. 18, l. 25-28; p. 19, l. 8-21

ISSUE PRESENTED FOR REVIEW

1. Whether Claims 1-8 were properly rejected under 35 U.S.C. §101 as directed to non-statutory matter.

2. Whether Claims 1-8 were properly rejected under 35 U.S.C. §112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which appellant regards as the invention.

3. Whether Claims 1-5 and 8 were properly rejected under 35 U.S.C. §101(b) and (e)(2) as anticipated by U.S. Patent 5,353,236 to Subbiah.

GROUPING OF THE CLAIMS

Appellants do not believe that any special grouping of the claims leads to a better understanding of the issues.

ARGUMENT

Appellant respectfully traverses the rejection of the claims under 35 U.S.C. §101 as directed to non-statutory subject matter. The Examiner has rejected Claims 1-4 under 35 U.S.C. §101, remarking that the claimed process is directed to non-statutory

subject matter since the process manipulates electron density data "without resulting in any physical transformation outside of a computer or other computational device." As noted in MPEP 2106.IV.B.2.(b).(i), a process is clearly statutory "if it requires physical acts to be performed outside the computer . . . . But, "[i]f a claim does not clearly fall into one or both of the safe harbors, the claim may still be statutory if it is limited to a practical application in the technological arts." The next section of MPEP provides an example: " . . .a computer process that simply calculates a mathematical algorithm that models noise is nonstatutory. However, a claimed process for digitally filtering noise employing a mathematical algorithm is statutory."

> The notion of "physical transformation" can be misunderstood. In the first place, it is not an invariable requirement, but merely one example of how a mathematical algorithm may bring about a useful application.
> ***AT&T Corp. v. Excel Communications, Inc.***, 172 F.3d 1352, 50 USPQ 2d 1447, 1454 (Fed. Cir. 1999), *cert denied*, 120 S. Ct. 368 (1999), *on remand*, 52 USPQ2d 1865 (D. Del. 1999)
>
> Today, we hold that the transformation of data, representing discrete dollar amounts, by a machine through a series of mathematical calculations into a final share price, constitutes a practical application of a mathematical algorithm, formula, or calculation, because it produces "a useful, concrete and tangible result"--a final share price momentarily fixed for recording and reporting purposes and even accepted and relied upon by regulatory authorities and in subsequent trades.
> ***State Street Bank & Trust Co. v. Signature Fin. Group, Inc.***, 47 USPQ 2d 1596, 1601 (Fed. Cir.), ***cert. denied***, 525 U.S. 1093 (1999)
>
> It is clear from the written description of the . . . patent that AT&T is only claiming a process that uses the Boolean principle in order to determine the value of the PIC indicator. The PIC indicator represents information about the call recipient's PIC, a useful, non-abstract result that facilitates differential billing of long-distance calls made by an IXC's subscriber. Because the claimed process applies the Boolean principle to produce a use, concrete, tangible result without pre-empting other uses of the mathematical principle on its face the claims process comfortably falls within the scope of Section 101. *See Arrhythimia Research Tech. Inc. v. Corazonix Corp.*, 958 R.2d 1053, 1060, 22 USPQ2d 1033, 1039 (Fed. Cir. 1992) ('That the product is numerical is not a criterion of whether the claim is directed to statutory subject.') *Id.*.
> ***AT&T Corp. v. Excel Communications, Inc.***, *supra.* at 1452.

Appellant's claimed method is the application of mathematical algorithms to modify "an electron density map of an experimental crystal structure," resulting in a new

electron density map, as recited in Claim 10. There is no longer in the law any requirement that the method result in any "physical transformation" as would be required by the Examiner. Further, the application of the recited mathematical manipulations is clearly directed a specified application, the formation of a revised electron density map of a crystal structure from a starting electron density map. There is no attempt to claim or forestall the use of any mathematical manipulation in any other application. See, e.g., the following claim steps:

(a) obtaining by x-ray diffraction observed structure factor amplitudes for a plurality of reflection from the crystal structure;

(b) selecting a starting set of crystallographic phases . . .;

(d) identifying features of the first electron density map . . .;

(e) making a comparison between the first electron density map and the expected distribution of electron density;

(g) establishing crystallographic phase probability distributions from the comparisons . . .;

(i) deriving an updated electron density map using crystallographic phases determined to be most probable . . . .

Independent Claims 1-8 clearly produce a concrete, tangible result within the teachings of AT&T Corp., *supra.,* and State Street Bank & Trust Co., supra. Even assuming that the electron density map is "the formation of data based on a crystal structure," as characterized by the Examiner, this is not a criteria for determining whether the claims are directed to statutory subject matter.

Appellant respectfully traverses the rejection of Claims 1-8 under 35 U.S.C. §112, second paragraph, as being indefinite for reciting "a plurality of reflections." No specific number of reflections are claimed or taught in appellant's specification since persons of ordinary skill in the art select some number of reflections depending on a desired resolution, as illustrated in Subbiah at Col. 8, lines 1-9.

The Examiner does not question the use of the term "plurality" and comments that "A plurality of reflections is reasonably interpreted as being as few as two."

> In rejecting a claim under the second paragraph of 35 USC 112, it is incumbent on the examiner to establish that one of ordinary skill in the pertinent art, when reading the claims in light of the supporting specification, would not have been

> able to ascertain with a reasonable degree of precision and particularity the particular area set out and circumscribed by the claims.
> **Ex parte Wu**, 10 USPQ2d 2031, 2033 (B.P.A.I. 1989)
>
> An applicant is entitled to claims as broad as the prior art and his disclosure will allow.
> **In re Rasmussen**, 211 USPQ 323, 326 (C.C.P.A. 1981)

Appellant has distinctly claimed a plurality of reflections since at least two reflections are required to perform the process claimed by appellant. However, there is no upper limit on the number of reflections that might be used. Indeed, an electron density map can be constructed from a single reflection (see, e.g., Subbiah at Col. 4, lines 29-32) so that the claimed process could be practiced with as few as two reflections. The exact number of reflections will simply be determined to a resolution determined by the experimenter. Appellant's process provides a modified first electron density map by recognizing features in an initial map that yield expected electron density distributions, which are used to obtain crystallographic phase probability distributions. This is done for all of the plurality (at least two) of reflections, where the most probable crystallographic phases are selected from the resulting maps to provide an updated electron density map. No undue experimentation is required for this determination since a large number of reflections are conventionally recorded, as illustrated by Subbiah.

The rejection of Claims 1-8 under 35 U.S.C. §112, second paragraph, should not be sustained.

Finally, appellant respectfully traverses the rejections of Claims 1-5 and 8 under 35 U.S.C. §102(b) and (e)(2) as being clearly anticipated by U.S. Patent 5,353,236 to Subbiah. Subbiah begins with measured amplitudes of structure factors, but no phase information, and yields phases and an electron density map. See, e.g., Col. 4, lines 27-35:

> The process is started with a low-resolution envelope of the macromolecular crystal. That envelope is used to obtain the phrase of the structure factor for one (or a few) low-resolution reflections. The phase of that structure factor is then used to construct a new, higher resolution envelope which is, in turn, used to calculate the phase for a higher resolution reflection so that an even higher resolution envelope can be constructed.

In another aspect, Subbiah finds arrangements of atomic scatterers that lead to calculated amplitudes of structure factors that are maximally consistent with measured amplitudes of structure factors.

In contrast, the claimed process of the present invention begins with measured amplitudes of structure factors and a set of starting phases are selected, not calculated from an envelope, and yields estimates of phases and an electron density map that have reduced bias. The input phases are adjusted to yield a map that has characteristics anticipated from the map features, but that were not used in constructing the initial estimates of phases. Appendix B presents a comparison of appellant's claim limitations with the Examiner's remarks and the corresponding teachings of Subbiah to the extent appellant could determine which claim limitation was covered by a reference to Subbiah.

To anticipate appellant's claimed invention, Subbiah must disclose every limitation in appellant's claimed process.

> We think the precise language of 35 U.S.C 102 that "a person shall be entitled to a patent unless," concerning novelty and unobviousness, clearly places a burden of proof on the Patent Office which requires it to produce the factual basis for its rejection of an application under sections 102 and 103 . . . .
> **In re Warner**, 154 USPQ 173, 177 (C.C.P.A. 1967, **cert. denied**, 389 U.S. 1057 (1968).
>
> An anticipating reference must describe the patented subject matter with sufficient clarity and detail to establish that the subject matter existed and that its existence was recognized by persons of ordinary skill in the field of the invention.
> **ATD Corp. v. Lyndall, Inc.**, 48 USPQ2d 1321, 1328 (Fed. Cir. 1998).

Referring to Appendix B, it is clear that Subbiah fails to disclose at least the following claimed process steps:

(b) selecting a starting set of crystallographic phases to combine with the observed structure factor amplitudes to form a first set of structure factors;

(d) identifying features of the first electron density map to obtain expected distributions of electron density;

(e) making a comparison between the first electron density map and the expected distribution of electron density;

(f) estimating how changes in the crystallographic phase of a reflection *k* affect the comparison;

(g) establishing crystallographic phase probability distributions from the comparisons for the possible crystallographic phases of reflection *k*;

(h) repeating steps (c) through (g) as *k* is indexed through all of the plurality of reflections;

(i) deriving an updated electron density map using crystallographic phases determined to be most probable from the crystallographic phase probability distributions for each one of the reflections;

(j) repeating steps (d) through (i) to obtain a final set of crystallographic phases with minimum bias from known electron density maps.

Subbiah, Col. 10, line 48, through Col. 21, line 38, referenced by the Examiner to show details of the Subbiah improvement process, teaches only moving scatterers about the map grid, calculating the Fourier amplitudes as the scatterers are moved, and correlating the calculated amplitudes with experimental X-ray diffraction data. A person skilled in the art would not possibly recognize Subbiah as having any teaching about establishing comparisons by altering crystallographic phases to establish crystallographic phase probability distributions.

The rejection of Claims 1-8 under 35 U.S.C. §102(b) and (e)(2) should not be sustained.

CONCLUSION

Appellants believe that the Examiner has not made a *prima facie* case for the rejections of currently pending Claims 1-8 under 35 U.S.C. §101, 35 U.S.C. §112, second paragraph, or 35 U.S.C. §102(b) and (e)(2). Appellants have definitely described and claimed a statutory process that is not taught by Subbiah. The rejection of Claims 1-8 should be reversed and this case passed to issue.

Respectfully submitted,

Date: 9-13-04

Signature of Attorney

Reg. No. 28,351
Phone (505) 665-3112

Ray G. Wilson
Los Alamos National Laboratory
LC/IP, MS A187
Los Alamos, New Mexico 87545

APPENDIX A - CLAIMS ON APPEAL

1. A method for improving an electron density map representing a crystal structure comprising:

(a) obtaining by x-ray diffraction observed structure factor amplitudes for a plurality of reflections from the crystal structure;

(b) selecting a starting set of crystallographic phases to combine with the observed structure factor amplitudes to form a first set of structure factors;

(c) deriving a first electron density map from the first set of structure factors;

(d) identifying features of the first electron density map to obtain expected distributions of electron density;

(e) making a comparison between the first electron density map and the expected distribution of electron density;

(f) estimating how changes in the crystallographic phase of a reflection *k* affect the comparison;

(g) establishing crystallographic phase probability distributions from the comparisons for the possible crystallographic phases of reflection *k*;

(h) repeating steps (c) through (g) as *k* is indexed through all of the plurality of reflections;

(i) deriving an updated electron density map using crystallographic phases determined to be most probable from the crystallographic phase probability distributions for each one of the reflections;

(j) repeating steps (d) through (i) to obtain a final set of crystallographic phases with minimum bias from known electron density maps; and

(k) forming a final electron density map using the final set of crystallographic phases.

2. The method of Claim 1, wherein identifying features of the electron density map includes making probability estimates of whether each point in the map is located in a solvent region or a crystal structure region.

3. The method of Claim 1, wherein identifying features of the election density map includes estimates of whether the electron density at each point in the map is

related by non-crystallographic symmetry to electron density at another point in the map.

4. The method of Claim 1, includes estimates of whether a structural motif is located at each point in the map.

5. The method of Claim 4, wherein the structural motif is a helix.

6. The method of any one of Claims 1, 2, 3, or 4, wherein the crystallographic phase probability distributions are log-likelihood functions.

7. The method of Claim 1, further including the steps of calculating first and second derivatives for the crystallographic phase probability distributions with respect to the structure factors; and

applying an FFT-based algorithm to determine the most probable crystallographic phase probability distributions.

8. The method of Claim 1, wherein the step of selecting a starting set of crystallographic phases includes;

selecting a model crystal structure having similarities to the crystal structure being examined;

assigning a low weighting factor to structure factors of the model crystal structure; and

combining the weighted structure factors with the observed structure factors for deriving the first electron density map.

APPENDIX B
CLAIM COMPARISON WITH REJECTION

Claim limitation	Examiner's comment	Reference citation	Appellant's comment
1. A method for improving an electron density map representing a crystal structure comprising:	Subbiah is directed to the crystallographic modeling of macromolecules as cited in the title and abstract with the construction of regions of units cells from diffraction patterns and Fourier amplitudes and to calculate electron density distributions as is also the subject matter of the instant claims.	Abstract: A method for constructing an image of a macromolecular crystal includes steps of providing an envelope which defines the region of a unit cell occupied by the macromolecule; distributing a collection of scattering bodies within the envelope; condensing the collection of scattering bodies to an arrangement that maximized the correlation between the diffraction pattern of the crystal and a pattern of Fourier amplitudes for the collection of scattering bodies; determining the phase associated with at least one of the Fourier amplitudes of the condensed collection of scattering bodies; calculating an electron density distribution of the crystal from the phase information; and defining an	Appellant's invention is directed to providing an improved electron density map of a crystal structure. Subbiah teaches a method for obtaining a high resolution of an envelope of the crystal structure (column 21, lines 23-26).

		image of the macro molecule in the electron density distribution.	
(a) obtaining by x-ray diffraction observed structure factor amplitudes for a plurality of reflections from the crystal structure;	This column 21 citation (column 21, lines 5-16) also discloses the utilization of the reflections in the diffraction pattern as also instantly claimed.	After 100-200 reflections have been used to calculate new envelopes, it will often be desirable to step in larger increments (i.e., more than one reflection will be phased in a given PW ["phase walk"] step). This will expedite the procedure, often without introducing significant new error. In addition, any such new errors are likely to be due to the weaker reflections. Thus, the risk of introducing error at these larger PW steps can be minimized by considering only the stronger reflections. Preferably, the larger PW steps will be done in increments of up to about 15% of the total number reflections phased thus far.	Subbiah does use reflection in a diffraction pattern, as does the present invention, but the initial calculations are used to obtain the phase of the structure factor (column 10, lines 58-61), not structure factor amplitudes.
(b) selecting a starting set of crystallographic phases to combine with the observed structure factor amplitudes to form a first set of structure factors;	In column 4, lines 22-42, the construction of a low resolution envelope for the electron density distribution is disclosed	The present invention produces a high-resolution model of the electron density distribution of a macromolecule in a defined asymmetric unit of a crystal lattice. This is accomplished	In Subbiah, the phases are determined from the reflection data, as stated in the citation. Applicant selects a starting set of crystallographic

		through a simple and rapid method for determining the phases of the reflection data for the macromolecule of interest. The process is started with a low-resolution envelope of the macromolecular crystal. That envelope is used to obtain the phase of the structure factor for one (or a few) low-resolution reflection. The phase of that structure factor is then used to construct a new, higher resolution, envelope which is, in turn, used to calculate the phase so a higher resolution reflection so that an even higher resolution envelope can be constructed. In this manner, the resolution of the envelope is improved by bootstrapping the solution from earlier calculations and the diffraction data. The process can be terminated at any stage, regardless of resolution. Thus, if the desired resolution is only intermediate, the process of this invention can be	phases to combine with observed structure factors amplitudes derived from the reflection data to form a first set of structure factors. This starting set of crystallographic phases is selected from a model or other source (page 5, lines 19-22), not the reflection data.

		terminated after the diffraction data of intermediate resolution has been phased.	
(c) deriving a first electron density map from the first set of structure factors;		Column 11, lines 22-26: If the low-resolution image is provided in the form of an electron density map, it [the image] can be expanded by simply choosing its boundaries to be the region circumscribed by a relatively low electron density contour.	In Subbiah, the electron density map is used only as an image to establish the envelope that is progressively refined by Subbiah. No further use is made of the electron density map.
(d) identifying features of the first electron density map to obtain expected distributions of electron density;	Particular structural motifs as in instant claims 4 and 5 are recognized in the map in the reference as disclosed in column 21, lines 34-39. Solvent regions and corresponding probability estimates are also described in the reference in column 20, lines 36-41, as required in instant claim 2.	For proteins, structural motifs such as inter-domain clefts and other prominent surface indentations, are typically observed at low resolution. At higher resolution, sheets, helices, side chains, and ultimately, atoms may be observed. As an example, the solvent may be expected to occupy 55% of the asymmetric unit volume (and the macro-molecule would occupy the remaining 45%). The scatterers might initially be placed in many more grip elements than would be expected for the	Subbiah uses the high resolution envelope to display features of the macromolecule, as particularly discussed at column 21, lines 23-39. There is no teaching about using the features to obtain any expected distributions of electron density. The example of Subbiah simply expresses a percentage of the cell unit volume occupied by solvent and by macromolecule, not any distribution of electron density.

		macromolecule.	
(e) making a comparison between the first electron density map and the expected distribution of electron density;	It is noted that each envelope of higher resolution is an estimated electron density distribution which is then compared with further phase refinement and reflection calculations to result in such practice as in the instant claims. (No specific citation was provided, so see column, lines 3-12)	After the new electron density map has been prepared, scatterers are placed in regions of high electron density. Typically, the asymmetric unit is divided into a grid of perpendicular lines, defining boxes that can each accommodate a single scatterer. As the resolution increases in succeeding PW steps, the fineness of the grid should also increase to allow for additional scatterers per unit volume. The grid will preferably accommodate three scatterers (and generally in the range of 1 to 6) per one-dimensional unit of the current resolution.	Subbiah makes no comparison between an electron density map and an expected distribution of electron density.
(f) estimating how changes in the crystallographic phase of a reflection *k* affect the comparison;			There is no citation to a comparable step in Subbiah.
(g) establishing crystallographic phase probability distributions from the comparisons for the			There is no citation to a comparable step in Subbiah.

possible crystallographic phases of reflection *k*;			
(h) repeating steps (c) through (g) as *k* is indexed through all of the plurality of reflections;			Subbiah does use a plurality of reflection in the process to increase the resolution of the macromolecule envelope developed by Subbiah.
(i) deriving an updated electron density map using crystallographic phases determined to be most probable from the crystallographic phase probability distributions for each one of the reflections;	This low-resolution envelope (column 4, lines 22-42) for electron density is then improved by the phase thereof being utilized for the construction of new higher resolution, envelopes in an comparative and iterative process for the electron density distribution as being modeled for the macromolecule. This improvement is detailed further in column 19, line 48, through column 21, line 38, wherein the desired resolution is obtained.	See above for column 4, lines 22-42. [The lengthy section of column 19, line 48, through column 21, line 38 is not reproduced. The text is discussed in the argument.]	There is no teaching in Subbiah about crystallographic phase probability distributions to determine the most likely phase for use in an updated electron density map.
(j) repeating steps (d) through (i) to obtain a final set of crystallographic phases with minimum bias from known electron density maps; and	This resolution is the final set of crystallographic determined electron density distribution with the corresponding probable phases and minimum bias compared to the actual		

	macromolecule structure as required in instant claim 1. The errors are also minimized for new envelopes also as a minimum bias as in instant claim 1 as described in column 21, lines 5-16.		
(k) forming a final electron density map using the final set of crystallographic phases.			

APPENDIX C
U.S. Patent Application S.N. 09/512,962

LIKELIHOOD-BASED MODIFICATION OF EXPERIMENTAL CRYSTAL STRUCTURE ELECTRON DENSITY MAPS

Inventor: Thomas C. Terwilliger
68 Camino Espejo
Santa Fe, NM 87501

CITIZEN OF THE UNITED STATES

EXPRESS MAIL CERTIFICATE EJ425552853US

LIKELIHOOD-BASED MODIFICATION OF EXPERIMENTAL CRYSTAL STRUCTURE ELECTRON DENSITY MAPS

RELATED APPLICATIONS

This application claims the benefit of U.S. provisional patent application S.N. 60/135,252, filed May 21, 1999.

STATEMENT REGARDING FEDERAL RIGHTS

This invention was made with government support under Contract No. W-7405-ENG-36 awarded by the U.S. Department of Energy. The government has certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates generally to the determination of crystal structure from the analysis of diffraction patterns, and, more particularly, to macromolecular crystallography.

BACKGROUND OF THE INVENTION

The determination of macromolecular structures, e.g., proteins, by X-ray crystallography is a powerful tool for understanding the arrangement and function of such macromolecules. Very powerful experimental methods exist for determining crystallographic features, e.g., structure factors and phases. While the structure factor amplitudes can be determined quite well, it is frequently necessary to improve or extend the phases before a realistic atomic model of the macromolecule, such as an electron density map, can be built.

Many methods have been developed for improving the phases by modifying initial experimental electron density maps with prior knowledge of characteristics expected in these maps. The fundamental basis of density modification methods is

that there are many possible sets of structure factors (amplitudes and phases) that are all reasonably probable based on the limited experimental data that is obtained from a particular experiment, and those structure factors that lead to maps that are most consistent with both the experimental data and the prior knowledge are the most likely overall. In these methods, the choice of prior information that is to be used, and the procedure for combining prior information about electron density with experimentally-derived phase information are important features.

Until recently, electron density modification has generally been carried out in a two-step procedure that is iterated until convergence. In the first step, an electron density map obtained experimentally is modified in real space in order to make it consistent with expectations. The modification can consist of, e.g., flattening solvent regions, averaging non-crystallographic symmetry-related regions, or histogram-matching. In the second step, phases are calculated from the modified map and are combined with the experimental phases to form a new phase set.

The disadvantage of this real-space modification approach is that it is not at all clear how to weight the observed phases from those obtained from the modified map. This is because the modified map contains some of the same information as the original map and some new information. This has been recognized for a long time and a number of approaches have been designed to improve the relative weighting from these two sources, including the use of maximum-entropy methods, the use of weighting optimized using cross-validation, and "solvent-flipping."

A comprehensive theory of the phase problem in X-ray crystallography and a formalism for solving it based on maximum entropy and maximum likelihood methods has been presented by Bricogne, Acta Cryst. A40, pp. 410-445 (1984) and Bricogne, Acta Cryst. A44, pp. 517-545 (1988). This formalism describes the contents of a crystal in terms of a collection of point atoms along with probabilities for their positions. From the positions of these atoms, crystallographic structure factors can be calculated, with a certainty depending on the certainties of the positions of the atoms. Extensions of the formalism are described in Bricogne (1988). The extended formalism specifically addresses the situation encountered in

crystals of macromolecules in which defined solvent and macromolecule regions exist in the crystallographic unit cell, and formulas for calculating probabilities of structure factors based on the presence of "flat" solvent regions are presented (Bricogne, 1988). The implementation of this formalism is not straightforward according to Xiang et al., Acta Cryst. D49, pp. 193-212 (1993), who point out that a full fledged implementation of this approach would be highly desirable and would provide a statistical technique for enforcing solvent flatness in advance. Xiang et al (1993) report that they settled for an approximation in which solvent flatness outside the envelope is imposed after the calculation of a model for the distribution of atoms, which corresponds to the existing procedure of flattening the solvent in an electron density map (Wang, Methods Enzymol. 115, pp. 90-112 (1985)).

The present invention solves the same problem that earlier procedures proposed by Bricogne (1988) address, and also includes the use of likelihood as a basis for choosing optimal crystallographic structure factors. The assumptions used in the present procedure differ substantially from those used by Bricogne (1988). For treatment of solvent and macromolecule (protein) regions in a crystal, Bricogne develops statistical relationships among structure factors based on a model of the contents of the crystal in which point atoms are randomly located, but in which atoms in the protein region are sharply-defined with low thermal parameters and atoms in the solvent region are diffuse, with high thermal parameters. In the present approach, no assumptions about the presence of atoms or possible values of thermal factors are used. Instead, it is assumed that values of electron density in the protein and solvent regions, respectively, are distributed in the same way in the crystal as in a model calculation of a crystal that may or may not be composed of discrete atoms.

The methods used to find likely solutions to the phase problem are also very different in the present approach compared to that of Bricogne (1988) because the assumptions used require the problem to be set up in different ways. Bricogne (1988) applies a maximum-entropy formalism developed by Bricogne (1984) to find likely arrangements of atoms in the crystal, which in turn can be used to calculate

the arrangement of electron density in the crystal. In the present method, likely values of the structure factors are found by applying a likelihood-based approach based on a combination of experimental information and the likelihood of resulting electron density maps. These structure factors can be used to calculate an electron density map that is then, in turn, a likely arrangement of electron density in the crystal.

Various objects, advantages and novel features of the invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following or may be learned by practice of the invention. The objects and advantages of the invention may be realized and attained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

SUMMARY OF THE INVENTION

In accordance with the purposes of the present invention, as embodied and broadly described herein, the present invention includes a method for improving an electron density map of an experimental crystal structure. A likelihood of a set of structure factors $\{\mathbf{F_h}\}$ is formed for the experimental crystal structure as (1) the likelihood of having obtained an observed set of structure factors $\{\mathbf{F_h}^{OBS}\}$ if structure factor set $\{\mathbf{F_h}\}$ was correct, and (2) the likelihood that an electron density map resulting from $\{\mathbf{F_h}\}$ is consistent with selected prior knowledge about the experimental crystal structure. The set of structure factors $\{\mathbf{F_h}\}$ is then adjusted to maximize the likelihood of $\{\mathbf{F_h}\}$ for the experimental crystal structure.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of the specification, illustrate embodiments of the present invention and, together with the description, serve to explain the principles of the invention. In the drawings:

FIGURE 1 is a flow sheet for a process to obtain characteristics from a model electron density map.

FIGURE 2 is a flow sheet for a process to derive structure factors consistent with experimental results which result in an electron density map with expected characteristics.

FIGURE 3A is a computer-generated electron density map provided by SOLVE software and calculated using only one substituted selenium atom.

FIGURE 3B is a computer-generated model electron density map calculated from an atomic model of the selected protein.

FIGURE 3C is a computer-generated electron density map derived from the process shown in FIGURES 1 and 2.

FIGURE 3D is a computer-generated electron density map derived from alternate available software called “dm”.

DETAILED DESCRIPTION

In accordance with the present invention, experimental phase information is combined with prior knowledge about expected electron density distribution in maps by maximizing a combined likelihood function. The fundamental idea is to express knowledge about the probability of a set of structure factors $\{\mathbf{F_h}\}$ ($\mathbf{F_h}$ includes amplitude , F_h, and phase, ϕ factors) and in terms of two quantities: (1) the likelihood of having measured the observed set of structure factors $\{\mathbf{F_h^{OBS}}\}$ if this structure factor set $\{\mathbf{F_h}\}$ were correct; and (2) the likelihood that the map resulting from this structure factor set $\{\mathbf{F_h}\}$ is consistent with prior knowledge about the structure under observation and other macromolecular structures. The index factor $\mathbf{h}$ is defined in terms of the hkl plane and unit vectors $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ in reciprocal lattice space as $\mathbf{h} = \mathrm{h}\mathbf{a}^* + \mathrm{k}\mathbf{b}^* + \mathrm{l}\mathbf{c}^*$.

When formulated in this manner, the overlap of information that occurred in the real-space modification methods is not present because the experimental and

prior information are kept separate. Consequently, proper weighting of experimental and prior information only requires estimates of probability functions for each source of information.

The likelihood-based density modification approach has a second very important advantage. This is that the derivatives of the likelihood functions with respect to individual structure factors can be readily calculated in reciprocal space by Fast Fourier Transform (FFT) based methods. As a consequence, density modification simply becomes an optimization of a combined likelihood function by adjustment of structure factors. This makes density modification a remarkably simple but powerful approach, requiring only that suitable likelihood functions be constructed for each aspect of prior knowledge that is to be incorporated.

The basic idea of the likelihood-based density modification procedure is that there are two key kinds of information about the structure factors for a crystal of a macromolecule. The first is the experimental phase and amplitude information, which can be expressed in terms of a likelihood (or a long-likelihood function $LL^{OBS}(\mathbf{F}_h)$ for each structure factor F_h. The experimental probability distribution for the structure factor, $p^{OBS}(\mathbf{F}_h)$ is given by

$$p^{OBS}(\mathbf{F}_h) = \exp\{LL^{OBS}(\mathbf{F}_H)\} \quad (1)$$

For reflections with accurately-measured amplitudes, the chief uncertainly in F_h will be in the phase, while for unmeasured or poorly-measured reflections, it will be in both phase and amplitude.

The second kind of information about structure factors in this formulation is the likelihood of the map resulting from the factors. For example, for most macromolecular crystals, a set of structure factors $\{\mathbf{F}_h\}$ that leads to a map with a flat region corresponding to solvent is more likely to be correct than one that leads to a map with uniform variation everywhere. This map likelihood function describes the probability that the map obtained from a set of structure factors is compatible with expectations:

$$p^{MAP}(\mathbf{F}_h) = \exp\{LL^{MAP}(\mathbf{F}_H)\} \quad (2)$$

The two principal sources of information are then combined, along with any prior knowledge of the structure factors, to yield the likelihood of a particular set of structure factors:

$$LL(\{\mathbf{F}_h\}) = LL^0(\{\mathbf{F}_h\}) + LL^{OBS}(\{\mathbf{F}_h\}) + LL^{MAP}(\{\mathbf{F}_h\}) \quad (3)$$

where $LL^0(\{\mathbf{F}_h\})$ includes any structure factor information that is known in advance, such as the distribution of intensities of structure factors.

In order to maximize the overall likelihood function in Eq. (3), the change in the map likelihood function in response to changes in structure factors must be known. In the case of the map likelihood function, $LL^{MAP}(\{\mathbf{F}_h\})$, there are two linked relationships: the response of the likelihood function to changes in electron density, and the changes in electron density as a function of changes in structure factors. In principle, the likelihood of a particular map is a complicated function of the electron density over the entire map. Furthermore, the value of any structure factor affects the electron density everywhere in the map.

For simplification, a low-order approximation to the likelihood function for a map is used instead of attempting to evaluate the function precisely. As Fourier transformation is a linear process, each reflection contributes independently to the electron density at a given point in the cell. Although the log-likelihood of the electron density might have any form, it is expected that for sufficiently small changes in structure factors, a first-order approximation to the log-likelihood function would apply and each reflection would also contribute relatively independently to changes in the log-likelihood function.

Consequently, a local approximation to the map likelihood function can be constructed, neglecting correlations among different points in the map and between reflections, expecting that it might describe with reasonable accuracy how the likelihood function would vary in response to small changes in the structure factors. By neglecting correlations among different points in the map, the log-likelihood for

the whole electron density map is written as the sum of the log-likelihood of the densities at each point in the map, normalized to the volume of the unit cell and the number of reflections used to construct it:

$$LL^{MAP}\left(\{\mathbf{F}_\mathbf{h}\}\right) \approx \frac{N_{REF}}{V}\int_V LL\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}\}\right)d^3\mathbf{x} \tag{4}$$

where N_{REF} is the number of independent reflections and V is the volume.

By treating each reflection as independently contributing to the likelihood function, a local approximation to the log-likelihood of the density at each point $LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}^0\}\right)\right)$ is written. This approximation is given by the sum over all reflections of the first few terms of a Taylor's series expansion around the value obtained with the starting structure factors $\{\mathbf{F}_\mathbf{h}^0\}$ used in a cycle of density modification,

$$LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}\}\right)\right) \approx LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}^0\}\right)\right) + \tag{5}$$

$$+\sum_\mathbf{h}\left[\Delta F_{\mathbf{h},\parallel}\frac{\partial}{\partial F_{\mathbf{h},\parallel}}LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}\}\right)\right)+\frac{1}{2}\Delta F_{\mathbf{h},\parallel}^2\frac{\partial^2}{\partial F_{\mathbf{h},\parallel}^2}LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}\}\right)\right)+\right.$$

$$\left.\Delta F_{\mathbf{h},\perp}\frac{\partial}{\partial F_{\mathbf{h},\perp}}LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}\}\right)\right)+\frac{1}{2}\Delta F_{\mathbf{h},\perp}^2\frac{\partial^2}{\partial F_{\mathbf{h},\perp}^2}LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_\mathbf{h}\}\right)\right)+\ldots\right],$$

where $\Delta F_{\mathbf{h},\parallel}$ and $\Delta F_{\mathbf{h},\perp}$ are the differences between $\mathbf{F}_\mathbf{h}$ and $\mathbf{F}_\mathbf{h}^0$ along the directions $\mathbf{F}_\mathbf{h}^0$ and $i\mathbf{F}_\mathbf{h}^0$, respectively.

Combining Eqs. (4) and (5) results in an expression for the map log-likelihood function,

$$LL^{MAP}(\{\mathbf{F_h}\}) \approx LL^{MAP}\left(\rho\left(\mathbf{x},\{\mathbf{F_h^0}\}\right)\right) + \frac{N_{REF}}{V}\sum_{\mathbf{h}}\left[\Delta F_{\mathbf{h},\parallel}\int_V \frac{\partial}{\partial F_{\mathbf{h},\parallel}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right)d^3\mathbf{x} + \frac{1}{2}\Delta F_{\mathbf{h},\parallel}^2\int_V \frac{\partial^2}{\partial F_{\mathbf{h},\parallel}^2} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right)d^3\mathbf{x} + \Delta F_{\mathbf{h},\perp}\int_V \frac{\partial}{\partial F_{\mathbf{h},\perp}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right)d^3\mathbf{x} + \frac{1}{2}\Delta F_{\mathbf{h},\perp}^2\int_V \frac{\partial^2}{\partial F_{\mathbf{h},\perp}^2} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right)d^3\mathbf{x} + \ldots\right] \quad (6)$$

The integrals in Eq. (6) can be rewritten in a form that is suitable for evaluation by a FFT-based approach. Considering the first integral in Eq. (6), use the chain rule to write,

$$\frac{\partial}{\partial F_{\mathbf{h},\parallel}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right) = \frac{\partial}{\partial \rho(\mathbf{x})} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right)\frac{\partial}{\partial F_{\mathbf{h},\parallel}}\rho(\mathbf{x}) \quad (7)$$

and note that the derivative of $\rho(\mathbf{x})$ with respect to $F_{\mathbf{h},\parallel}$ for a particular index value $\mathbf{h}$ is given by,

$$\frac{\partial}{\partial F_{\mathbf{h},\parallel}}\rho(\mathbf{x}) = \frac{2}{V}\mathrm{Re}\left[e^{i\phi_{\mathbf{h}}-2\pi i\mathbf{h}\cdot\mathbf{x}}\right] \quad (8)$$

Now the first integral in Eq. (6) is rewritten in the form,

$$\int_V \frac{\partial}{\partial F_{\mathbf{h},\parallel}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right)d^3\mathbf{x} = \frac{2}{V}\mathrm{Re}\left[e^{i\phi_{\mathbf{h}}}\mathrm{a}_{\mathbf{h}}^*\right] \quad (9)$$

where the complex number $\mathrm{a}_{\mathbf{h}}$ is a term in the Fourier transform of

$$\frac{\partial}{\partial \rho(\mathbf{x})} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right),$$

$$\mathrm{a}_{\mathbf{h}} = \int_V \frac{\partial}{\partial \rho(\mathbf{x})} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right)e^{2\pi i\mathbf{h}\cdot\mathbf{x}}d^3\mathbf{x} \quad (10)$$

In space groups other than P1, only a unique set of structure factors needs to be specified to calculate an electron density map. Taking space group symmetry into account, Eq. (9) can be generalized to read,

$$\int_V \frac{\partial}{\partial F_{\mathbf{h},\parallel}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right) d^3\mathbf{x} = \frac{2}{V}\sum_{\mathbf{h}'} \mathrm{Re}\left[e^{i\phi_{\mathbf{h}'}} \mathbf{a}^*_{\mathbf{h}'}\right] \tag{11}$$

where the indices $\mathbf{h}'$ are all indices equivalent to $\mathbf{h}$ due to space-group symmetry.

A similar procedure is used to rewrite the second integral in Eq. (6), yielding the expression,

$$\int_V \frac{\partial^2}{\partial F^2_{\mathbf{h},\parallel}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right) d^3\mathbf{x} = \frac{2}{V^2}\sum_{\mathbf{h}',\mathbf{k}'} \mathrm{Re}\left[e^{-i\phi_{\mathbf{h}'}} e^{i\phi_{\mathbf{k}'}} \mathbf{b}_{\mathbf{h}'-\mathbf{k}'} + e^{-i\phi_{\mathbf{h}'}} e^{-i\phi_{\mathbf{k}'}} \mathbf{b}_{\mathbf{h}'+\mathbf{k}'}\right] \tag{12}$$

where the indices $\mathbf{h}'$ and $\mathbf{k}'$ are each all indices equivalent to $\mathbf{h}$ due to space group symmetry, and where the coefficients $b_{\mathbf{h}}$ are again terms in a Fourier transform, this time the second derivative of the log-likelihood of the electron density,

$$b_{\mathbf{h}} = \int_V \frac{\partial^2}{\partial^2 \rho(\mathbf{x})^2} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right) e^{2\pi i \mathbf{h}\cdot\mathbf{x}} d^3\mathbf{x} \tag{13}$$

The third and fourth integrals in Eq. (6) can be rewritten in a similar way yielding the expressions,

$$\int_V \frac{\partial}{\partial F_{\mathbf{h},\perp}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right) d^3\mathbf{x} = \frac{2}{V}\sum_{\mathbf{h}'} \mathrm{Re}\left[e^{i\phi_{\mathbf{h}'}} \mathbf{a}^*_{\mathbf{h}'}\right] \tag{14}$$

and

$$\int_V \frac{\partial^2}{\partial F^2_{\mathbf{h},\perp}} LL\left(\rho\left(\mathbf{x},\{\mathbf{F_h}\}\right)\right) d^3\mathbf{x} = \frac{2}{V^2}\sum_{\mathbf{h}',\mathbf{k}'} \mathrm{Re}\left[e^{-i\phi_{\mathbf{h}'}} e^{i\phi_{\mathbf{k}'}} \mathbf{b}_{\mathbf{h}'-\mathbf{k}'} - e^{-i\phi_{\mathbf{h}'}} e^{-i\phi_{\mathbf{k}'}} \mathbf{b}_{\mathbf{h}'+\mathbf{k}'}\right] \tag{15}$$

The significance of Eqs. (4) through (15) is that there is now a simple expression (Eq. (6)) describing how the map likelihood function $LL^{MAP}(\{\mathbf{F_h}\})$ varies when small changes are made in the structure factors. Evaluating this expression requires only that the first and second derivatives of the log-likelihood of the electron density be calculated with respect to electron density at each point in the map (see Eq. (22) below) and that a Fast Fourier Transform (FFT) be carried out as described by Teneyck, Acta Cryst. 33, pp. 486-492 (1977), incorporated by reference. Furthermore, maximization of the (local) overall likelihood function (Eq.

(3)) becomes straightforward, as every reflection is treated independently. It consists simply of adjusting each structure factor to maximize its contribution to the approximation to the likelihood function through Eqs. (3)-(15).

In practice, instead of directly maximizing the overall likelihood function, it is used here to estimate the probability distribution for each structure factor, and then to integrate this probability distribution over the phase (or phase and amplitude) of the reflection to obtain a weighted mean estimate of the structure factor. Using Eqs. (3)-(15), the probability distribution for an individual structure factor can be written as,

$$\begin{aligned} \ln p(\mathbf{F}_\mathbf{h}) \approx {} & LL^0(\mathbf{F}_\mathbf{h}) + LL^{OBS}(\mathbf{F}_\mathbf{h}) + \\ & \frac{2N_{REF}}{V^2}\Delta F_{\mathbf{h},\parallel} \sum_{\mathbf{h}'} \mathrm{Re}\left[e^{i\phi_{\mathbf{h}'}} a^*_{\mathbf{h}'}\right] + \\ \frac{2N_{REF}}{V^3}\Delta F_{\mathbf{h},\parallel}\, 2\sum_{\mathbf{h}',\mathbf{k}'} \mathrm{Re}\left[e^{-i\phi_{\mathbf{h}'}} e^{i\phi_{\mathbf{k}'}} b_{\mathbf{h}'-\mathbf{k}'} + e^{-i\phi_{\mathbf{h}'}} e^{-i\phi_{\mathbf{k}'}} b_{\mathbf{h}'+\mathbf{k}'}\right] + & \\ & \frac{2N_{REF}}{V^2}\Delta F_{\mathbf{h},\perp} \sum_{\mathbf{h}'} \mathrm{Re}\left[e^{i\phi_{\mathbf{h}'}} a^*_{\mathbf{h}'}\right] + \\ \frac{2N_{REF}}{V^3}\Delta F_{\mathbf{h},\perp}\, 2\sum_{\mathbf{h}',\mathbf{k}'} \mathrm{Re}\left[e^{-i\phi_{\mathbf{h}'}} e^{i\phi_{\mathbf{k}'}} b_{\mathbf{h}'-\mathbf{k}'} - e^{-i\phi_{\mathbf{h}'}} e^{-i\phi_{\mathbf{k}'}} b_{\mathbf{h}'+\mathbf{k}'}\right] & \end{aligned} \tag{16}$$

where, as above, the indices $\mathbf{h}'$ and $\mathbf{k}'$ are each all indices equivalent to $\mathbf{h}$ due to space group symmetry, and the coefficients $a_\mathbf{h}$ and $b_\mathbf{h}$ are given in Eqs. (10) and (13). Also, as before, $\Delta F_{\mathbf{h},\parallel}$ and $\Delta F_{\mathbf{h},\perp}$ are the differences between $\mathbf{F}_\mathbf{h}$ and $\mathbf{F}^0_\mathbf{h}$ along the directions $\mathbf{F}^0_\mathbf{h}$ and $i\mathbf{F}^0_\mathbf{h}$, respectively. All the quantities in Eq. (16) can be readily calculated once a likelihood function for the electron density and its derivatives are obtained (see Eq. (22) below).

A key step in likelihood-based density modification is the decision as to the likelihood function for values of the electron density at a particular location in the map. For the present purposes, an expression for the log-likelihood of the electron density $LL\left(\rho(\mathbf{x},\{\mathbf{F}_\mathbf{h}\})\right)$ at a particular location $\mathbf{x}$ in a map is needed that depends on whether the point satisfies any of a wide variety of conditions, such as being in the protein or solvent region of the crystal, being at a certain location in a known

fragment of structure, or being at a certain distance from some other feature of the map. Information can be incorporated on the environment of $\mathbf{x}$ by writing the log-likelihood function as the log of the sum of conditional probabilities dependent on the environment of $\mathbf{x}$,

$$LL\left(\rho\left(\mathbf{x},\{\mathbf{F}_{\mathbf{h}}\}\right)\right)=\ln\left[p\left(\rho(\mathbf{x})|PROT\right)p_{PROT}(\mathbf{x})+p\left(\rho(\mathbf{x})|SOLV\right)p_{SOLV}(\mathbf{x})\right] \quad (17)$$

where $p_{PROT}(\mathbf{x})$ is the probability that $\mathbf{x}$ is in the protein region and $p\left(\rho(\mathbf{x})|PROT\right)$ is the conditional probability for $\rho(\mathbf{x})$ given that $\mathbf{x}$ is in the protein region, and $p_{SOLV}(\mathbf{x})$ and $p\left(\rho(\mathbf{x})|SOLV\right)$ are the corresponding quantities for the solvent region. The probability that x is the protein or solvent region is estimated by a modification, described in Terwilliger, Acta Cryst. D55, pp. 1863-1871 (1999), of the methods described in Wang, Methods Enzymol. 115, pp. 90-112 (1985), and Leslie, Proceedings of the Study Weekend organized by CCP4, pp. 25-32 (1988), incorporated herein by reference. If there were more than just solvent and protein regions that identified the environment of each point, then Eq. (17) could be modified to include those as well.

In developing Eqs. (3)-(15), the derivatives of the likelihood function for electron density were intended to represent how the likelihood function changed when small changes in one structure factor were made. Surprisingly, the likelihood function that is most appropriate for the present invention is not a globally correct one. Instead, it is a likelihood function that represents how the overall likelihood function varies in response to small changes in one structure factor, keeping all others constant. To see the difference, consider the electron density in the solvent region of a macromolecular crystal. In an idealized situation with all possible reflections included, the electron density might be exactly equal to a constant in this region. The goal in using Eq. (16) is to obtain the relative probabilities for each possible value of a particular unknown structure factor $\mathbf{F}_{\mathbf{h}}$. If all other structure factors were exact, then the globally correct likelihood function for the electron

density (zero unless the solvent region is perfectly flat) would correctly identify the correct value of the unknown structure factor.

Now suppose the phase information is imperfect. The solvent regions would have a significant amount of noise, and the electron density value is no longer a constant. If the globally correct likelihood function is used for the electron density, a zero probability would be assigned to any value of the structure factor that did not lead to an absolutely flat solvent region. This is clearly unreasonable, because all the other (incorrect) structure factors are contributing noise that exists regardless of the value of this structure factor.

This situation is very similar to the one encountered in structure refinement of macromolecular structures where there is a substantial deficiency in the model. The errors in all the other structure factors in the discussion correspond to the deficiency in the macromolecular model in the refinement case. The appropriate variance to use as a weighting factor in refinement includes the estimated model error as well as the error in measurement. Similarly, the appropriate likelihood function for electron density for use in the present method is one in which the overall uncertainty in the electron density due to all reflections other than the one being considered is included in the variance.

A likelihood function of this kind for the electron density can be developed using a model in which the electron density due to all reflections but one is treated as a random variable. See Terwilliger et al., Acta Cryst. D51, pp. 609-618 (1996), incorporated herein by reference. Suppose that the true value of the electron density at $\mathbf{x}$ was known and was given by ρ_T. Then consider that there are estimates of all the structure factors, but that substantial errors exist in each one. The expected value of the estimate of this electron density (ρ_{OBS}) obtained from current estimates of all the structure factors will be given approximately by $<\rho_{OBS}>=\beta\rho_T$, and the expected value of the variance by $<(\rho_{OBS}-\beta\rho_T)^2>=\sigma^2_{MAP}$. The factor β represents the expectation that the calculated value of ρ will be smaller than the true value. This is true for two reasons. One is that such an

estimate may be calculated using figure-of-merit weighted estimates of structure factors, which will be smaller than the correct ones. The other is that phase error in the structure factors systematically leads to a bias towards a smaller component of the structure factor along the direction of the true structure factor.

A probability function for the electron density at a point $\mathbf{x}$ that is appropriate for assessing the probabilities of values of the structure factor for one reflection can now be written as,

$$p(\rho) = \exp - \frac{(\rho - \beta\rho_T)^2}{2\sigma_{MAP}^2} \tag{18}$$

In a slightly more complicated case where the value of ρ_T is not known exactly, but rather has an uncertainly σ_T, Eq. (18) becomes,

$$p(\rho) = \exp - \frac{(\rho - \beta\rho_T)^2}{2(\beta^2\sigma_T^2 + \sigma_{MAP}^2)} \tag{19}$$

Finally, in the case where only a probability distribution $p(\rho_T)$ for ρ_T is known, Eq. (18) becomes,

$$p(\rho) = \int_{\rho_T} p(\rho_T) \exp\left\{ - \frac{(\rho - \beta\rho_T)^2}{2\sigma_{MAP}^2} \right\} d\rho_T \tag{20}$$

Using Eqs. (19) and (20), a histogram-based approach (Goldstein et al., Acta Cryst. D54, pp. 1230-1244 (1998)) can be used to develop likelihood functions for the solvent region of a map and for the macromolecule-containing region of a map. The approach is simple. The probability distribution for true electron density in the solvent or macromolecule regions of a crystal structure is obtained from an analysis of model structures and represented as a sum of gaussian functions of the form,

$$p(\rho_T) = \sum_k w_k \exp\left\{ - \frac{(\rho - c_k)^2}{2\sigma_k^2} \right\} \tag{21}$$

where the coefficients w_k are normalized so that the integral of $p(\rho_T)$ is normalized over all ρ.

The coefficients c_k, σ_k^2, and w_k are obtained as follows. A model of a protein structure is used to calculate theoretical structure factors for a crystal of that protein structure. Exemplary structures may be obtained from the Protein Data Bank (H.M.Berman et al., The Protein Data Bank. Nucleic Acids Research 28, pp. 235-242, 2000), and containing space group, cell dimensions and angles, and a list of coordinates, atom types, occupancies, and atomic displacement parameters. The model may be chosen to be similar in size, resolution of the data, and overall atomic displacement factors to the experimental protein structure to be analyzed, but this is not essential to the process. The resolution of the calculated data and the average atomic displacement parameter may be adjusted to match those of the protein structure to be analyzed. Alternatively, a standardized resolution such as 3 Angstrom units and unadjusted atomic displacement parameters may be used, as in the examples given below. The theoretical structure factors for the model are then used to calculate an electron density map.

The electron density map is then divided into "protein" and "solvent" regions in the following way. All points in the map within a specified distance (typically 2.5 Angstrom units) of an atom in the model are designated "protein" and all others are designated "solvent". The next steps are carried out separately for "protein" and "solvent" regions of the electron density map. A histogram of the numbers of points in the protein or solvent region of the electron density map falling into each possible range of electron densities is calculated. The histogram is then normalized so that the sum of all histogram values is equal to unity. Finally, the coefficients c_k, σ_k^2, and w_k are obtained by least-squares fitting of Equation (21) to the normalized histograms. One set of coefficients is obtained for the "protein" region, another for the "solvent" region.

If the values of β and σ_{MAP} are known for an experimental map with unknown errors, but identified solvent and protein regions, the probability distribution for electron density in each region of the map can be written approximately from Eq. (19) as,

$$p(\rho) = \sum_k w_k \exp\left\{-\frac{(\rho - \beta c_k)^2}{2\left(\beta^2 \sigma_k^2 + \sigma_{MAP}^2\right)}\right\} \quad (22)$$

with the appropriate values of β and σ_{MAP} and separate values of c_k, σ_k^2, and w_k for protein and solvent regions. In practice, the values of β and σ_{MAP} are estimated by a least-squares fitting of the probability distributions for protein and solvent regions given in Eq. (22) to the ones found in the protein and solvent regions in the experimental map.

This fitting is carried out by first constructing separate histograms of values of electron density in the protein and solvent regions defined by the methods described in Wang, Methods Enzymol 115, pp. 90-112 (1985) and Leslie, Proceedings of the Study Weekend, organized by CCP4, pp. 25-32 (1988), incorporated by reference. Next, the histograms are normalized so that the sum, over all values of electron density, of the values in each histogram is unity. In this way the histograms represent the probability that each value of electron density is observed. Then the values of β and σ_{MAP} in Eq. (22) are adjusted to minimize the squared difference between the values of the probabilities calculated from Eq. (22) and the observed values from the analysis of the histogram. This procedure has the advantage that the scale of the experimental map does not have to be accurately determined. Then Eq. (22) is used with the refined values of β and σ_{MAP} as the probability function for electron density in the corresponding region (solvent or macromolecule) of the map.

The process discussed above is more particularly shown in Figures 1 and 2. The basic process of maximum-likelihood density modification has two parts. In the first part, the characteristics of model electron density map(s) are obtained (Figure 1). These will typically be the same or similar for many different applications of the algorithm. In the second part (Figure 2), a particular set of structure factors has typically been obtained using experimental measurements on a crystal. This set of structure factors can be directly used to calculate an electron density map. Due to

uncertainties in measurement, the electron density map is imperfect. In this second part, a set of structure factors (phases and amplitudes) is found that is consistent with experimental measurements of those structure factors, and that, when used to calculate an electron density map, lead to an electron density that has characteristics similar to those obtained from the model electron density map(s). A likelihood-based approach is used to find this optimal set of structure factors.

Figure 1 shows a process for obtaining characteristics from model electron density maps to use in the above equations. First, a model protein structure obtained by X-ray crystallography is chosen 10. The model is used to conventionally calculate an electron density map 12. The electron density map is segmented into "protein" and "solvent" regions 14, where the protein region contains all points within a selected proximity to an atom in the model. Histograms of electron density are obtained 16 for "protein" and "solvent" regions. For protein and solvent regions, coefficients for the Gaussian function formed by Eq. (21) are found so that Eq. (21) is optimally fitted 18 to the histogram for that region. Eq. (21), with the fitted coefficients, is output 22 as the analytical description of the electron density distribution in the protein or solvent region for this model structure.

Figure 2 depicts the process for finding the optimal set of structure factors for a crystal consistent with experimental measurements and resulting in an the electron density map having characteristics expected from the model structure. The inputs are (1) the analytical descriptions of electron density distributions (Eq. 21) for model solvent and protein regions output from the process shown in Figure 1; (2) the fraction f_{solvent} of the crystal that is in the "solvent" region; (3) the space group and cell parameters of the crystal; and (4) the experimental measurements of structure factors (phases and amplitudes) and their associated uncertainties.

The overall process steps for estimating the probability that the electron density at each point in the map is correct are: (1) obtaining probability distributions for electron density for the protein and solvent regions of the current electron density map; (2) estimating the probability that the electron density at each point in

the map is correct; (3) evaluating how the probabilities would change if the electron density at each point in the map changed; (4) using a Fourier Transform to evaluate how the overall likelihood of the electron density map would change if one crystallographic structure factor changed; (5) combining the likelihood of the map with the likelihood of having observed the experimental data, as a function of each crystallographic structure factor; and (6) deriving a new probability distribution for each crystallographic structure factor. Steps (1) through (6) are then iterated until no substantial further changes in structure factors are obtained.

The process for finding structure factors that are consistent with experiments and that result in an electron density map with expected characteristics is shown in Figure 2. . The current best estimates of structure factors are used to calculate 32 an electron density map. If there is uncertainty in amplitude or phase, the weighted mean structure factor is ordinarily used, where all possible amplitudes and phases are weighted by their relative probabilities. The electron density map is segmented into protein and solvent regions as described by Wang, Methods Enzymol. 115, pp.90-112 (1985) and Leslie, Proceedings of the Study Weekend organized by CCP4, p. 25-32 (1988), incorporated by reference. The analytical descriptions of electron density distributions for model protein and solvent regions are fitted by least-squares to the observed electron density distributions in the protein and solvent regions in this electron density map using the factors β and σ^2_{map}, where the same values of β and σ^2_{map} are used for both protein and solvent regions.

Eq. (22), with the values of coefficients $c_k, \sigma^2_k,$ and w_k for protein and solvent regions obtained from fitting Eq. (21) to the model electron density from the process shown in Figure 1, and with the values of β and σ^2_{map} obtained above, now is an analytical description of a probability distribution for electron density in protein or solvent regions of the electron density map. The derivatives of Eq. (22) with respect to electron density (ρ) are obtained by standard procedures.

The probability of the electron density at each point in the protein or solvent regions of the current map is obtained 34 from Eq. (22). The logarithm of the overall log-likelihood of this map is calculated from the sum of the logarithms of these probabilities. The first and second derivatives with respect to electron density of the probability distributions for each point are calculated 36 to evaluate how the probability at each point would change if the electron density at each point in the map were changed.

An FFT is used to calculate 38, for each structure factor, how the overall log-likelihood of the map would change if that structure factor were changed. Then, the log-likelihood of the map as a function of all possible values of each structure factor is estimated 42 from a Taylor's series expansion of the log-likelihood of the map. This provides a log-likelihood estimate of any value of each structure factor as the sum of the log-likelihood of the resulting map with the log-likelihood of having observed the experimental data given that value.

The new estimate 44 of the logarithm of the probability that a structure factor has a particular value is obtained by adding together the log-likelihood of the map for that value of the structure factor and the log-likelihood of observing the experimental value of the structure factor. The exponentiation of these values is the probability of each possible value of a structure factor and is used to obtain a new weighted estimate of the structure factor. The new estimate of the structure factor is then returned to step 32 to begin a new iteration with a revised electron density map.

To evaluate the utility of maximum-likelihood density modification as described here, the process was applied to both model and real data. The first set of tests consisted of a set of phases constructed from a model with 32%-68% of the volume of the unit cell taken up by protein. The cell was in space group P21212 with cell dimensions of $a = 94, b = 80, c = 43$ Å and one molecule in the asymmetric unit, and was based on 6906 model data from ∞ to 3.0 Å calculated from coordinates from a dehalogenase enzyme from Rhodococcus species ATCC 55388

(ATCC, 1992), except that some of the atoms were not included to vary the fraction of solvent in the unit cell. Phases with simulated errors were generated by adding phase errors to yield an average value of the cosine of the phase error (i.e., the true figure of merit of the phasing) of, $<\cos(\Delta\phi)>$=0.42 for acentric and 0.39 for centric reflections.

Analyses were done using conventional real-space solvent flattening and reciprocal-space solvent flattening, Terwilliger, Acta Cryst. **D55**, pp. 1863-1871 (1999), incorporated by reference, as well as the maximum-likelihood approach. Both real-space and reciprocal-space solvent flattening improved the quality of phasing considerably. The real space density modification included both solvent flattening and histogram matching to be as comparable as possible to the maximum-likelihood density modification according to the present invention.

Table I shows the quality of phases obtained after each method for density

TABLE I

Fraction Protein (%)	Starting $<\cos(\Delta\phi)>$	Real Space $<\cos(\Delta\phi)>$	Reciprocal Space $<\cos(\Delta\phi)>$	Maximum Likelihood $<\cos(\Delta\phi)>$
32	.41	.64	.85	.87
42	.40	.62	.67	.83
50	.41	.54	.56	.77
68	.42	.48	.41	.53

modification was applied to this model case. In all cases, maximum-likelihood density modification of this map resulted in phases with an effective figure of merit ($<\cos(\Delta\phi)>$) higher than any of the other methods. When the fraction of solvent in the model unit cell was 50%, for example, maximum-likelihood density modification yielded an effective figure of merit of 0.83, while real-space solvent flattening and histogram matching resulted in an effective figure of merit of 0.62 and reciprocal-space solvent flattening yielded 0.67.

The utility of maximum-likelihood density modification was also compared with real-space density modification and with reciprocal-space solvent flattening

using experimental multiwavelength (MAD) data on initiation factor 5A (IF-5A). IF-5A crystallizes in space group I4 with cell dimensions of $a = 114, b = 114, c = 33$ Å, one molecule in the asymmetric unit, and a solvent content of about 60%. The structure was solved using MAD phasing based on three selenium atoms in the asymmetric unit at a resolution of 2.2 Å. For purposes of testing density modification methods, only one of the three selenium sites was used in phasing here, resulting in a starting map with a correlation coefficient to the map calcuclated using the final refined structure of 0.37.

Figures 3A-D show sections through electron density maps obtained after real-space density modification using solvent flattening and histogram matching and after maximum-liklihood density modification:

Figure 3A is an electron density map from SOLVE, calculated using only one substituted selenium atom;

Figure 3B is an electron density map determined from a model structure, calculated from an atomic model of the protein;

Figure 3C is an electron density map determined using the process of the present invention (RESOLVE);

Figure 3D is an electron density map calculated using a software program "dm," K. Cowtan, "dm: An automated procedure for phase improvement by density modification," Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31, pp. 34-38 (1994).

As anticipated, the "dm"-modified map is improved over the starting map and has a correlation coefficient of 0.65. The maximum-likelihood modified map is even more substantially improved with a correlation coefficient to the map based on a refined model of 0.79.

While the above demonstration considered only two sources of expected electron density distributions (probability distributions for solvent regions and for protein-containing regions), the methods can be applied directly to a wide variety of sources of information. For example, any source of information about the expected

electron density at a particular point in the unit cell that can be written in a form such as the one in Eq. (22) can be used in the procedure to describe the likelihood that a particular value of electron density is consistent with expectation.

Sources of expected electron density information that are especially suitable for application to the present method include non-crystallographic symmetry and the knowledge of the location of fragments of structure in the unit cell. In the case of non-crystallographic symmetry, the probability distribution for electron density at one point in the unit cell can be written using Eq. (22) with a value of ρ_T equal to the weighted mean at all non-crystallographically equivalent points in the cell. The value of σ_T can be calculated based on their variances and the value of σ_{MAP}. In the case of knowledge of locations of fragments in the unit cell, this knowledge can be used to calculate estimates of the electron density distribution for each point in the neighborhood of the fragment. These electron density distributions can then, in turn, be used as described above to estimate ρ_T and σ_T in this region.

An iterative process could be developed in which fragment locations are identified by cross-correlation or related searches, density modification is applied, and additional searches are carried out to further generate a model for the electron density. Such a process could potentially even be used to construct a complete probablistic model of a macromolecular structure using structure factor estimates obtained from molecular replacement with fragments of macromolecular structures as a starting point.

In all these cases, the electron density information could be included in much the same way as the probability distributions that are used herein for the solvent and protein regions of maps. In each case, the key is an estimate of the probability distribution for electron density at a point in the map that contains some information that restricts the likely values of electron density at that point. The procedure could be further extended by having probability distributions describing the likelihood that a particular point in the unit cell is within a protein region, within a solvent region, within a particular location in a fragment of protein structure, within a non-

crystallographically related region, and so on. These probability distributions could be overlapping or non-overlapping. Then, for each category of points, the probability distribution for electron density within that category could be formulated as in Eq. (22) and the method of the present invention applied.

This process extends reciprocal-space solvent flattening in two important ways. One is that the expected electron density distribution in the non-solvent region is included in the calculations, and a formalism for incorporating information about the electron density map from a wide variety of sources is developed. The second is that the probability distribution for the electron density is calculated using Eq. (22) for both solvent and non-solvent regions and values of the scaling parameter β and the map uncertainty σ_{MAP} are estimated by a fitting model and observed electron density distributions. This fitting process makes the whole procedure very robust with respect to scaling of the experimental data, which otherwise would have to be very accurate in order that the model electron density distributions be applicable.

The foregoing description of the invention has been presented for purposes of illustration and description and is not intended to be exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

WHAT IS CLAIMED IS:

1. A method for improving an electron density map of an experimental crystal structure, comprising the steps of:

a. forming a model electron density map of a model crystal structure;

b. forming model histograms of model electron densities in identified protein and solvent regions of the model electron density map;

c. fitting a model probability distribution function defined by

$$p(\rho_T) = \sum_k w_k \exp\left\{-\frac{(\rho - c_k)^2}{2\sigma_k^2}\right\}$$

to the model histograms, where k is separately indexed over the protein and solvent regions of the model map, $p(\rho_T)$ is the probability of an electron density at a point, w_k is a normalization factor, ρ is electron density, c_k is a mean value of ρ, and σ_k is a variance of ρ, where the fitting determines the coefficients w_k, c_k, and σ_k;

d. determining a set of experimental structure factors from x-ray diffraction data for the experimental crystal and forming an experimental electron density map;

e. forming separate experimental histograms of experimental electron densities over protein and solvent regions of the model electron density map;

f. fitting an experimental probability distribution function defined by

$$p(\rho_T) = \sum_k w_k \exp\left\{-\frac{(\rho - \beta c_k)^2}{2\left(\beta^2\sigma_k^2 + \sigma_{MAP}^2\right)}\right\}$$

to separate protein and solvent regions of the experimental histograms, where β is an expectation that an experimental value of ρ is less than a true value and σ_{map} is a variance, where the fitting determines the coefficients β and σ_{map};

g. determine from the probability distribution function the overall experimental log-likelihood of the electron density in the protein and solvent regions of the experimental map;

h. determine how the log-likelihood of the electron density of the protein and solvent regions of the experimental map would change as each experimental structure factor changes to output a revised log-likelihood of any value of each experimental structure factor; and

i. forming from the revised log-likelihood of experimental structure factor values a new set of structure factors and returning the new set of structure factors to step (f) to interate the process.

2. A method according to Claim 1, wherein step a. further includes the step of selecting the model crystal structure to be similar in size, data resolution, and atomic displacement factors to the experimental crystal.

3. A method according to Claim 1, wherein step b. further includes the step of identifying protein and solvent regions by designating all points within a selected distance of an atom as “protein” and all other points at “solvent.”

4. A method according to Claim 2, wherein step b. further includes the step of identifying protein and solvent regions by designating all points within a selected distance of an atom as “protein” and all other points at “solvent.”

5. A method according to Claim 1, wherein step h. includes the steps of forming a Taylor’s series expansion of the log-likelihood of the experimental map and evaluating terms of the Taylor’s series expansion using a Fast Fourier Transform.

6. A method for improving an electron density map of an experimental crystal structure, comprising the steps of:

a. forming a likelihood of a set of structure factors $\{\mathbf{F_h}\}$ for the experimental crystal structure as (1) the likelihood of having obtained an observed set of structure factors $\{\mathbf{F_h}^{OBS}\}$ if structure factor set $\{\mathbf{F_h}\}$ was correct, and (2) the likelihood that an electron density map resulting from $\{\mathbf{F_h}\}$ is consistent with selected prior knowledge about the experimental crystal structure; and

b. adjusting the set of structure factors $\{\mathbf{F_h}\}$ to maximize the likelihood of $\{\mathbf{F_h}\}$ for the experimental crystal structure.

7. A method according to Claim 6, wherein forming the likelihood of $\{\mathbf{F_h}\}$ further includes forming the likelihood that $\{\mathbf{F_h}\}$ is compatible with selected other prior knowledge of the experimental crystal structure.

8. A method according to Claim 6, wherein the step of adjusting the structure factors includes the steps of (1) determining the response of the likelihood of $\{\mathbf{F_h}\}$ to changes in the electron density map and (2) determining the response of the electron density map to changes in $\{\mathbf{F_h}\}$.

9. A method according to Claim 6, further including the step of approximating the likelihood of the electron density map includes the step of forming a Taylor's series expansion of the likelihood of the electron density map and evaluating the terms of the Taylor's series expansion through a Fast Fourier Transform.

ABSTRACT

A maximum-likelihood method for improves an electron density map of an experimental crystal structure. A likelihood of a set of structure factors $\{\mathbf{F_h}\}$ is formed for the experimental crystal structure as (1) the likelihood of having obtained an observed set of structure factors $\{\mathbf{F_h}^{OBS}\}$ if structure factor set $\{\mathbf{F_h}\}$ was correct, and (2) the likelihood that an electron density map resulting from $\{\mathbf{F_h}\}$ is consistent with selected prior knowledge about the experimental crystal structure. The set of structure factors $\{\mathbf{F_h}\}$ is then adjusted to maximize the likelihood of $\{\mathbf{F_h}\}$ for the experimental crystal structure. An improved electron density map is constructed with the maximized structure factors.

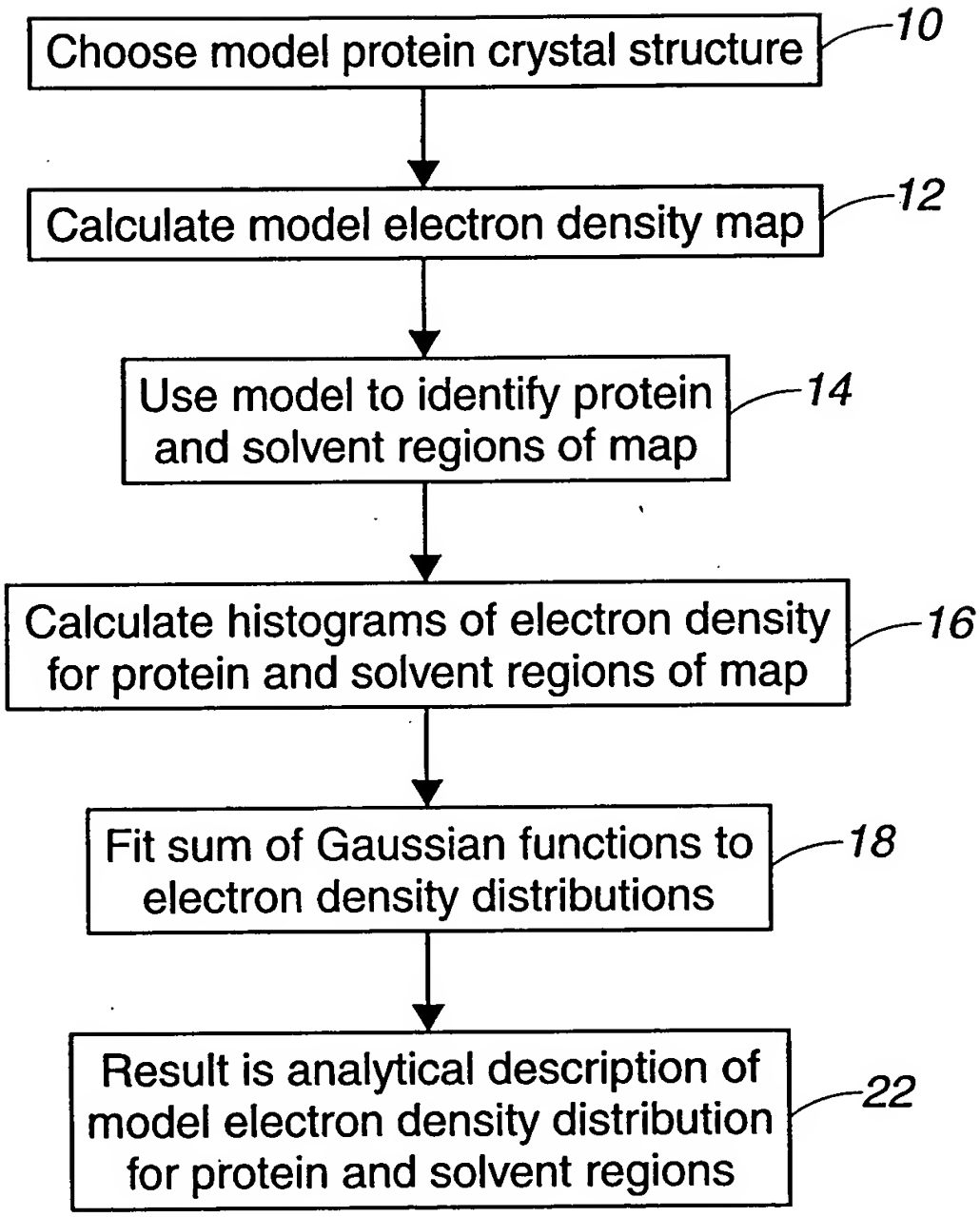
Choose model protein crystal structure
10
Calculate model electron density map
12
Use model to identify protein and solvent regions of map
14
Calculate histograms of electron density for protein and solvent regions of map
16
Fit sum of Gaussian functions to electron density distributions
18
Result is analytical description of model electron density distribution for protein and solvent regions
22

Fig. 1

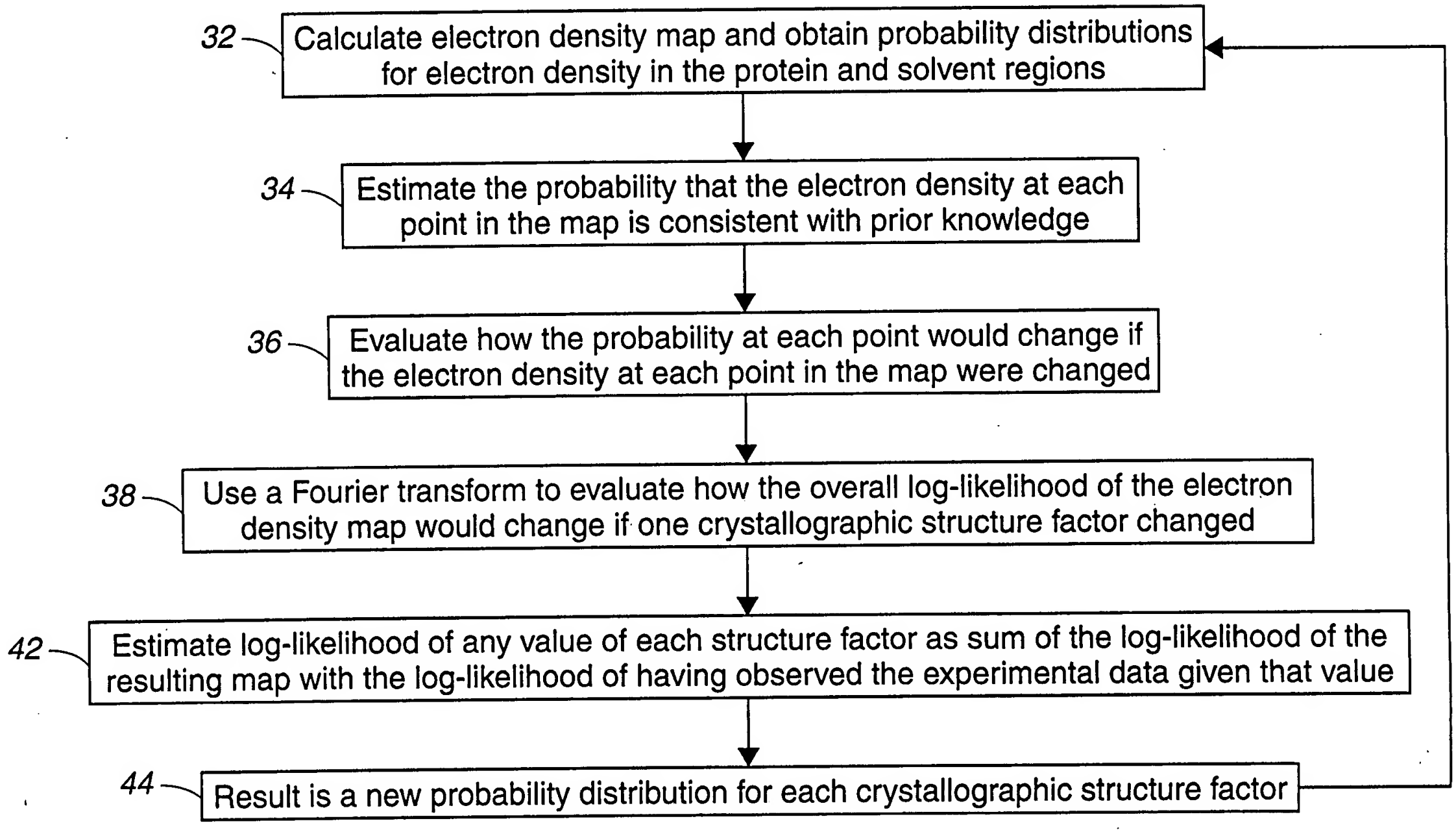
32
Calculate electron density map and obtain probability distributions for electron density in the protein and solvent regions
34
Estimate the probability that the electron density at each point in the map is consistent with prior knowledge
36
Evaluate how the probability at each point would change if the electron density at each point in the map were changed
38
Use a Fourier transform to evaluate how the overall log-likelihood of the electron density map would change if one crystallographic structure factor changed
42
Estimate log-likelihood of any value of each structure factor as sum of the log-likelihood of the resulting map with the log-likelihood of having observed the experimental data given that value
44
Result is a new probability distribution for each crystallographic structure factor

Fig. 2

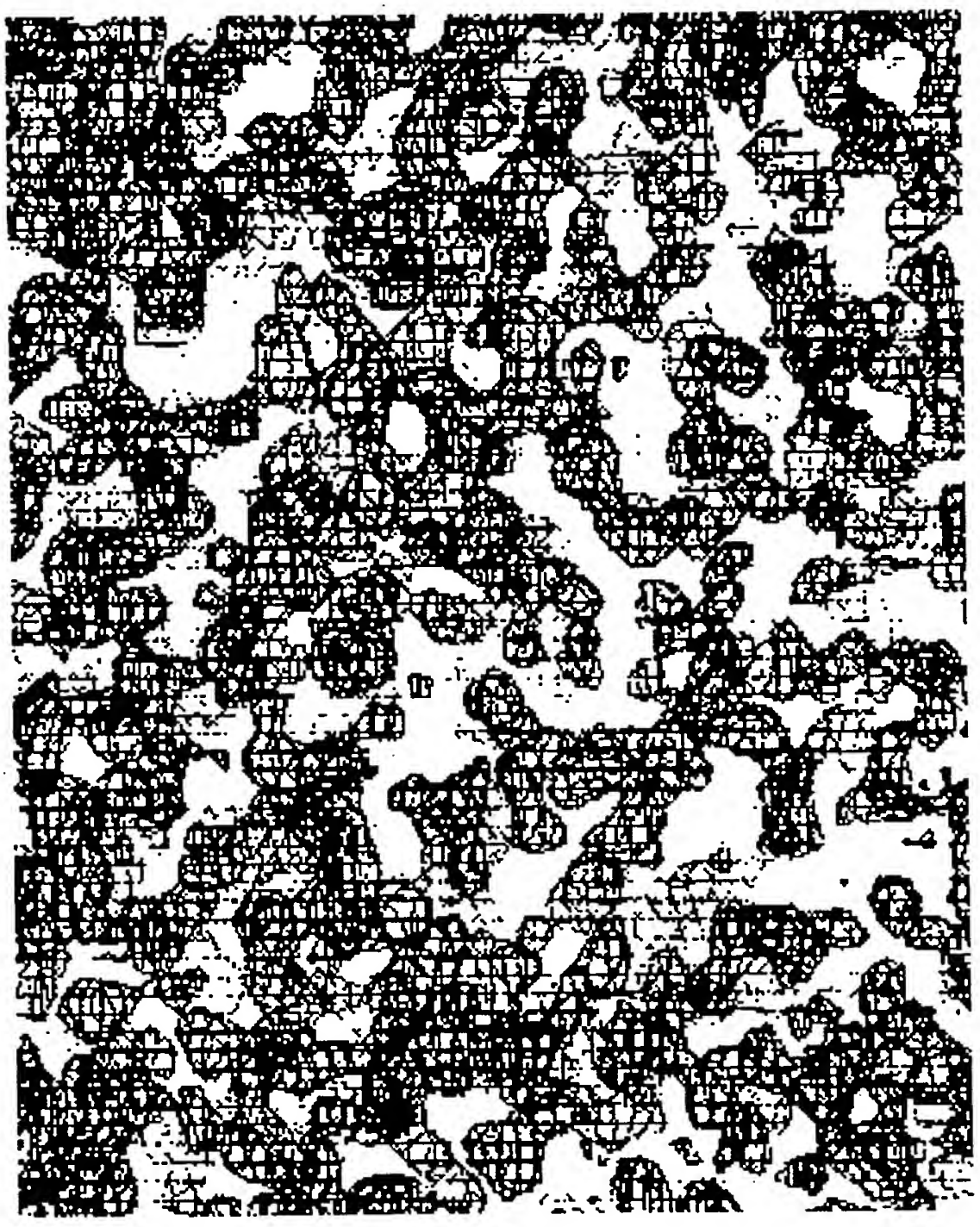

Fig. 3A

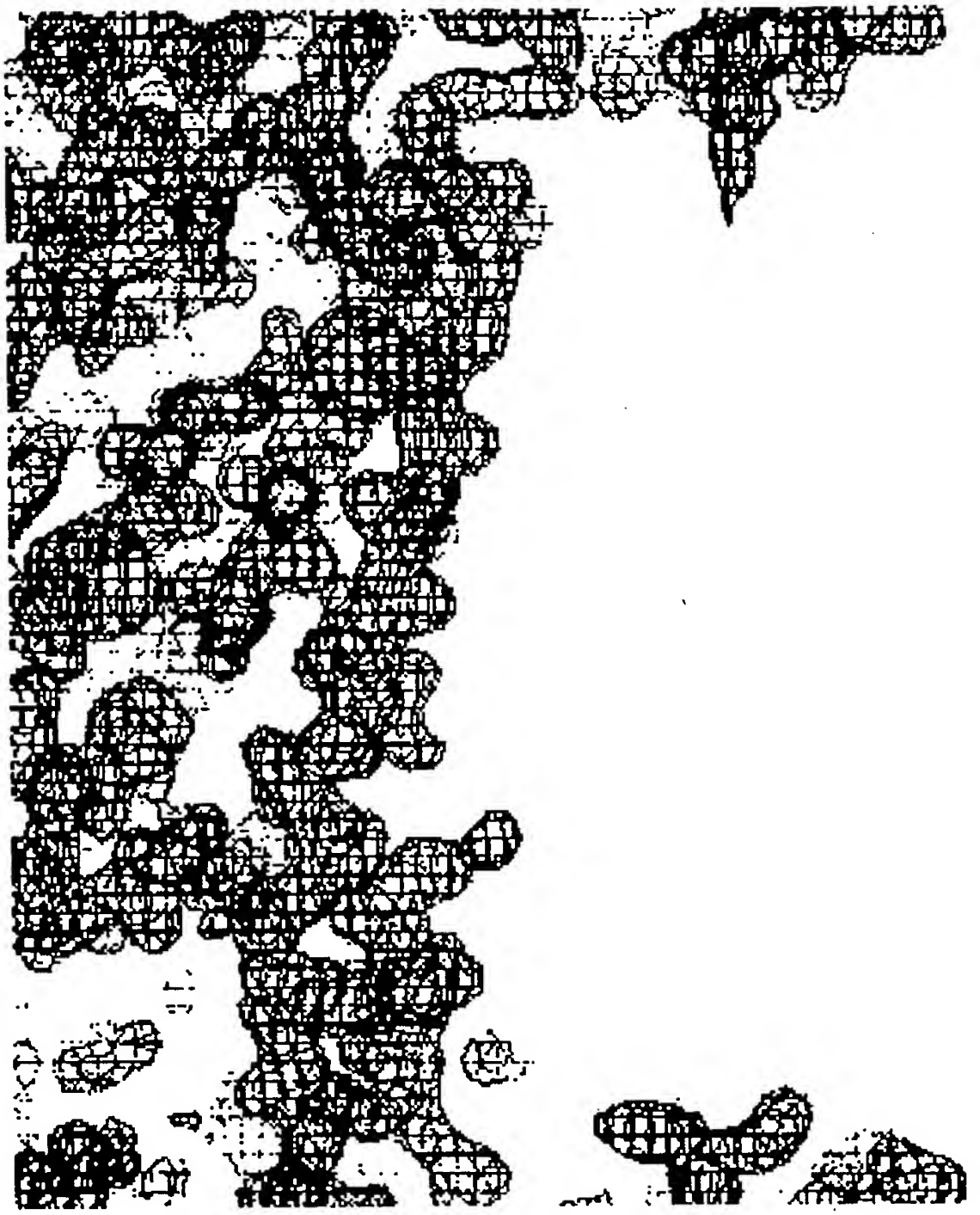

Fig. 3B

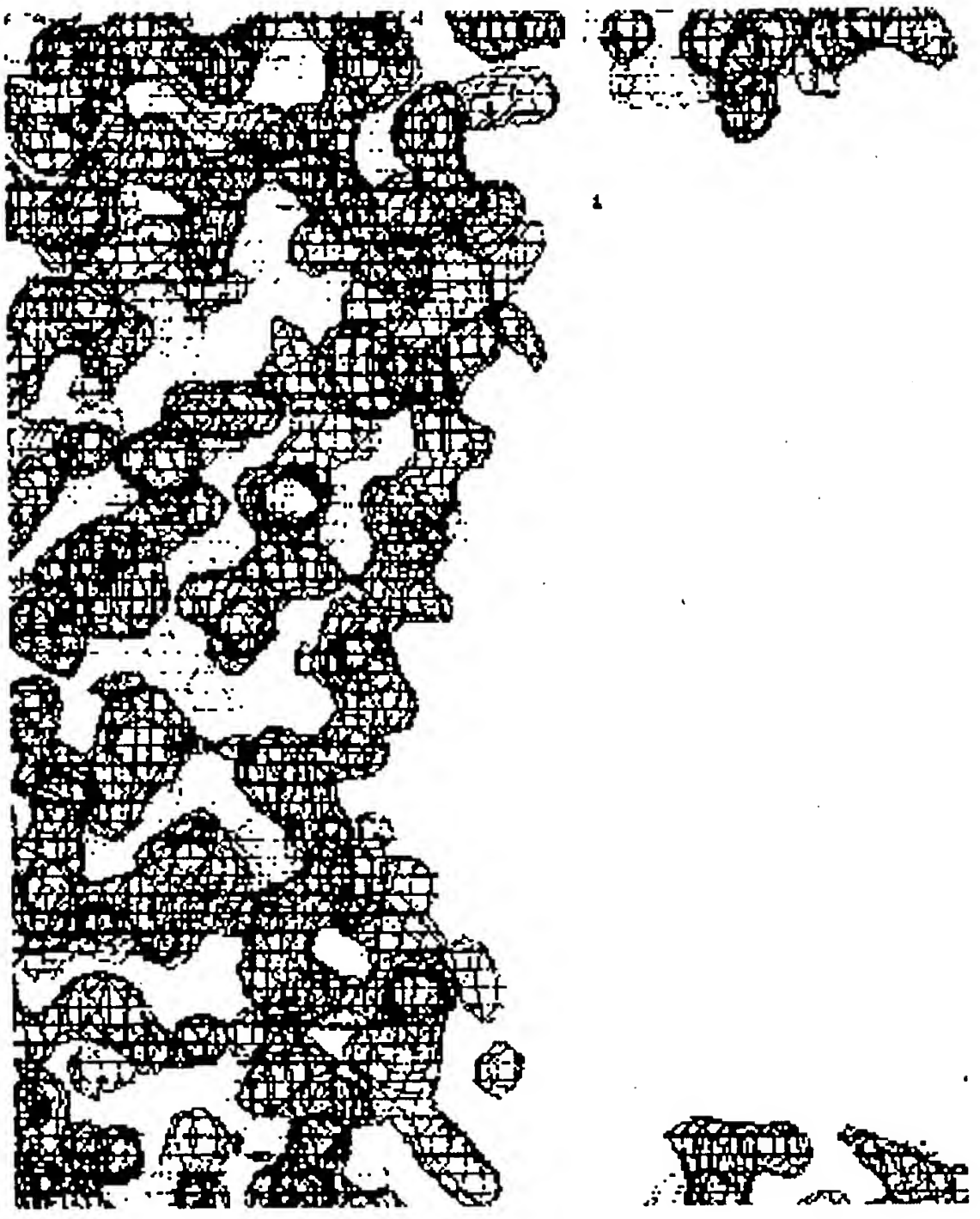

Fig. 3C

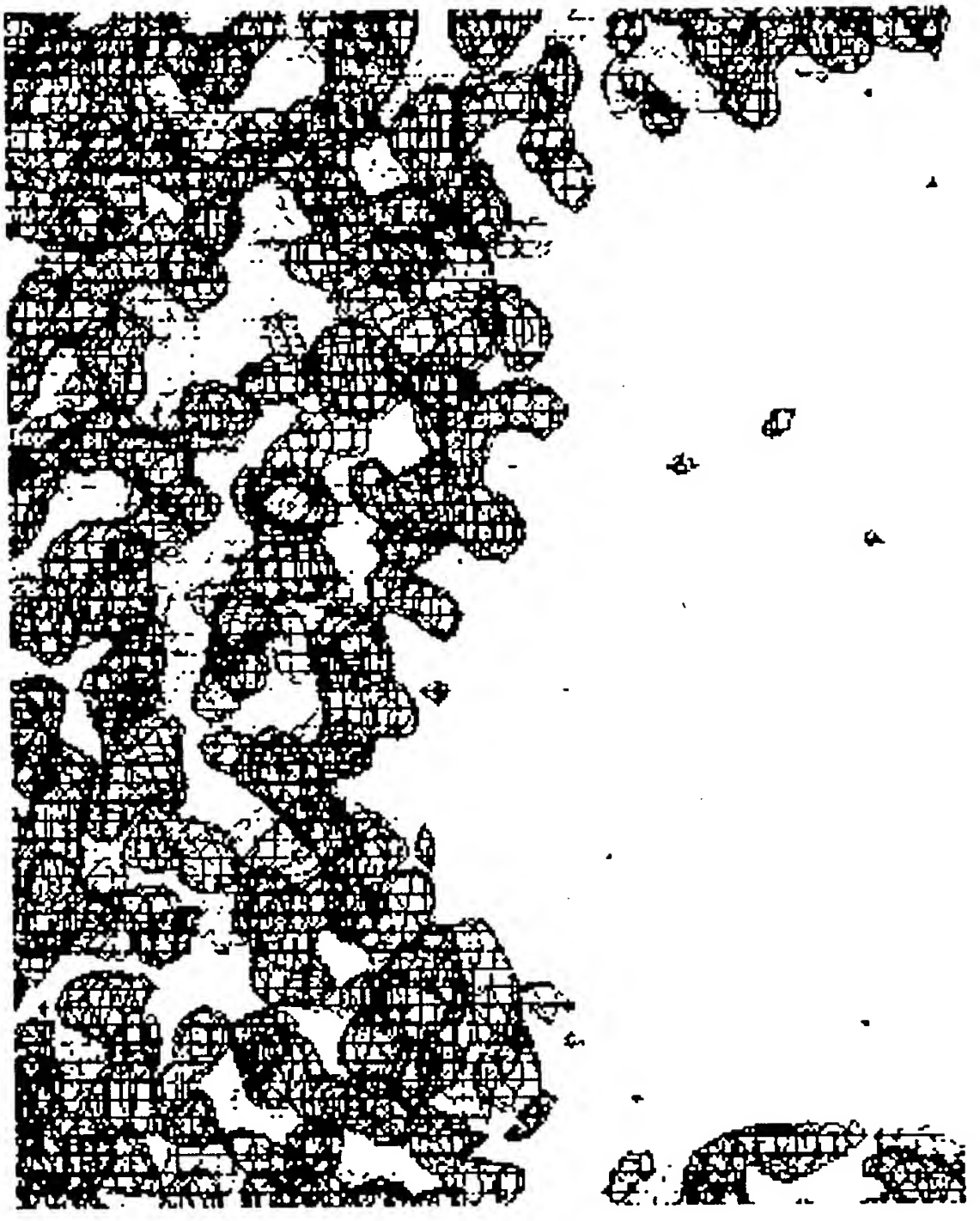

Fig. 3D